(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
30 August 2001 (30.08.2001)

PCT

(10) International Publication Number
WO 01/63937 A2

(51) International Patent Classification[7]: H04N 7/26

(21) International Application Number: PCT/US01/06094

(22) International Filing Date: 26 February 2001 (26.02.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
09/512,406 24 February 2000 (24.02.2000) US

(71) Applicant: SARNOFF CORPORATION [US/US]; 201 Washington Road, CN 5300, Princeton, NJ 08543-5300 (US).

(72) Inventor: REITMEIER, Glenn, Arthur; 193 Cinnabar Lane, Yardley, PA 19067 (US).

(74) Agents: MENDELSOHN, Steve et al.; Mendelsohn & Associates, P.C., 1515 Market Street, Suite 715, Philadelphia, PA 19102 (US).

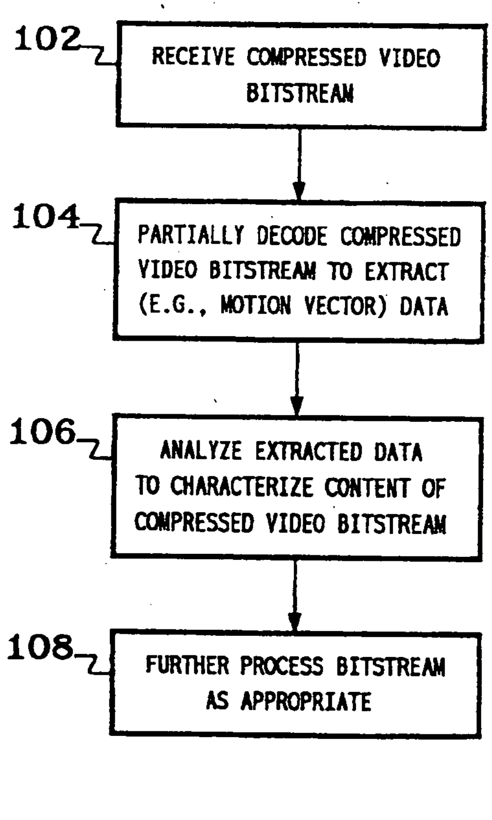
(81) Designated States *(national)*: BR, JP.

(84) Designated States *(regional)*: European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR).

Published:
— *without international search report and to be republished upon receipt of that report*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: COMPRESSED VIDEO ANALYSIS



(57) Abstract: A compressed digital video bitstream is partially decoded to extract certain data (e.g., inter/intra block field data, motion vector data, and/or DCT coefficient data). The extracted data is then analyzed (for a single frame or over multiple frames) to characterize picture content. For example, the inter/intra block field data can be analyzed to detect the existence of scene changes and camera switches. In addition to inter/intra block field data, patterns in the motion vector data can also be characterized to detect the existence of camera pans and zooms and the occurrence of persons/objects moving within or entering or leaving a scene. Intra/inter block field data and motion vector data can also be used to detect the occurrence of still text, slides, and pictures within a compressed bitstream. In addition, DCT coefficient data can be used to detect the existence of text and raster displays within the field of view.

WO 01/63937 A2

COMPRESSED VIDEO ANALYSIS

BACKGROUND OF THE INVENTION

Field of the Invention

The present invention relates to image processing, and, in particular, to the analysis of the content of a compressed video bitstream.

Description of the Related Art

Compressed digital video standards, such as H.261, MPEG-1, and MPEG-2, are on the verge of rapid deployment and proliferation in applications that include video teleconferencing, distributed multimedia systems, and television broadcasting. Unlike analog video signals, digital video signals employ many levels of data representation in order to effect their high compression rates. Typical operations that are performed in a video compression encoder include:

- o Organization of original image pixels as blocks of pixels;
- o Motion estimation during which blocks of pixels from surrounding frames are correlated with each block of the current frame to find a "best prediction," which is then encoded as a motion vector;
- o Motion compensation during which the residual interframe differences are generated between each current block and the corresponding "best prediction" block;
- o DCT transformation during which a discrete cosine transformation is applied to each block of residual interframe differences;
- o Quantization during which the resulting DCT transform coefficients are quantized;
- o Run-length encoding during which the resulting quantized DCT coefficients are run-length/value encoded;
- o Generation of a series of instructions that indicate the start of a block, the motion vector used to predict that block, the run/length value data for the quantized DCT coefficients of the residual interframe differences, and the end of the block; and
- o Variable-length encoding during which the instructions are variable-length coded (VLC) according to tables defined in the corresponding standard (e.g., based on Huffman coding principles) to form a serial compressed video bitstream.

In virtually every compressed digital video standard, only the VLC and instruction syntaxes are rigidly defined. This approach allows encoders to employ algorithms of varying cost/performance and allows encoder performance to improve over time, while maintaining full backwards compatibility with the installed base of decoders. However, because of the syntax elements that are part of a given compression standard, any reasonable encoder built to work with those standards will inevitably have similarities in certain algorithmic approaches, such as motion compensation.

In certain applications, it is desirable to determine information about the content of an existing compressed video bitstream. For example, it may be desirable to identify and catalog the locations of different scenes in a video stream in order to enable subsequent searching for specific scenes. One way to achieve this goal is to fully decode the compressed video bitstream to the decoded pixel domain to generate a corresponding decoded video stream, which can then be analyzed "manually" by a human operator or "automatically" using conventional analysis tools that identify the locations of transitions between scenes in the video stream.

SUMMARY OF THE INVENTION

The present invention is directed to techniques for analyzing the content of compressed video bitstreams without having first to fully decode the bitstream to the decoded pixel domain. According to the present invention, a compressed video bitstream is only partially decoded (e.g., just enough to extract the motion vector data) and the partially decoded data is then analyzed to characterize the content of the bitstream.

According to one embodiment, the present invention is a method for characterizing picture content of a compressed video bitstream, comprising the steps of: (a) partially decoding the compressed video bitstream to extract particular data for the compressed video bitstream; and (b) analyzing the extracted particular data to characterize the picture content of the compressed video bitstream.

BRIEF DESCRIPTION OF THE DRAWINGS

Other aspects, features, and advantages of the present invention will become more fully apparent from the following detailed description, the appended claims, and the accompanying drawings in which Fig. 1 is a flow diagram of processing, according to one embodiment of the present invention.

DETAILED DESCRIPTION

Compressed digital video signals can be analyzed to extract key high-level events that occur in their picture content, thus making it possible to monitor a compressed digital video bitstream for purposes such as cataloging, alerting, and/or key frame extraction. Processes can be developed to directly process the compressed bitstream to extract information such as (but not necessarily limited to):

(1) Scene changes;
(2) Camera switches;
(3) Camera pans;
(4) Camera zooms;
(5) Person/object moving within or entering/leaving a scene;
(6) Classifying occurrence of text, slide/picture, raster display, etc.; and
(7) Performing enhancement processing

Development of such approaches can enable a wide range of systems and applications as digital video becomes increasingly integrated into the information infrastructure.

With the combined knowledge of compression syntax, common encoder compression processing practices, and video signal processing expertise, a compressed digital video bitstream (such as a teleconferencing bitstream) can be a rich source of information. The compressed syntax-level representation contains clues to key high-level events that occur in the picture content.

Fig. 1 shows a block diagram of the processing of a compressed digital video bitstream, according to one embodiment of the present invention. The compressed digital video bitstream is received (step **102** in Fig. 1) and partially decoded (step **104**), for example, just enough to extract the motion vector data represented in the bitstream for each frame. Depending on the encoding scheme, this partial decoding may involve only the variable-length decoding of bitstream data and extraction of the motion vector data from the resulting variable-length decoded information. Most importantly, substantial computational advantage is gained by (a) avoiding the inverse DCT transform and (b) avoiding the recomputation of motion vectors or other data that is extractable from the compressed domain. Furthermore, storage requirements are significantly reduced by not having to store fully decoded pixel data

After extracting data for one or more frames, the extracted data is analyzed to characterize the content of the compressed digital video bitstream (step **106**). The type of analysis performed and the nature of the content characterized will vary from application to application. Some of these different applications are described below. Again, depending on the particular application, appropriate subsequent processing may be performed (step **108**) based on the characterized bitstream content. This subsequent processing may include cataloging the various scenes in the bitstream or any other suitable processing.

Many video compression standards rely on predictive (i.e., inter-frame) coding, where a frame is predicted from another (i.e., reference) frame in the video stream by computing a motion vector for each block of data that best predicts it from the reference frame. If there is no good predictor from the reference frame, a block of data may instead be intra-frame coded (i.e., encoded without reference to any other frame and therefore without a motion vector being assigned). According to embodiments of the present invention (as described below), information as to which blocks in the current frame have been encoded using intra-frame coding and which blocks have been encoded using inter-frame coding as well as the magnitudes and directions of the motion vectors used during the inter-frame coding may be analyzed to characterize the content of the compressed bitstream. In other embodiments, other information, such as the DCT coefficients, may be analyzed to characterize bitstream content.

<u>Frequency of Intra-Frame Coded Blocks</u>

As described above, during motion estimation processing, if a good predictor for a current block of pixel data cannot be found in the reference frame, the block may be encoded using intra-frame coding. A

given frame may be encoded with both intra-frame coded blocks and inter-frame coded blocks. The relative frequencies of intra-frame and inter-frame coded blocks per frame can be used to indicate certain types of changes in the picture content. In particular, if the number of intra-frame coded blocks in a current frame exceeds a specified threshold, then this may be an indication of the occurrence of a scene change or camera switch in the compressed bitstream.

Motion Vector Pattern Analysis

The locations, relative magnitudes, and directions of the set of motion vectors for a given frame can be used as an indication of temporal changes in the picture content of the compressed bitstream, especially when these patterns continue over multiple consecutive frames. Such motion vector pattern analysis can be used to distinguish different types of changes in picture content.

For example, during a camera pan, most of the picture content in the current frame was present in the previous frame, but at a uniformly translated location, with new information added to the field of view along one picture boundary (i.e., corresponding to the direction of the pan). As a result, the motion vectors for most of the current frame will have relatively uniform magnitude and direction, with the new information being represented either as intra-frame coded blocks or as inter-frame coded blocks with possibly uncorrelated motion vectors. Such a pattern of motion vectors and inter/intra block types can be used to detect the occurrence of a camera pan.

A camera "zoom in" may be detected as a set of motion vectors forming a radial pattern with the motion vectors generally referencing towards the focal point of the zoom. Similarly, a camera "zoom out" may be detected as a set of motion vectors forming a radial pattern with most of the motion vectors generally referencing away from the focal point of the zoom with a ring of intra-coded blocks and/or inter-coded blocks having uncorrelated motion vectors around the outer boundary of the frame corresponding to the new information added to the field of view during the camera zoom out.

Patterns within the motion vector field can also be used to indicate the motion of a person/object within a scene, or the entrance or exit of a person/object to or from a scene. A person/object moving within the camera's field of view will be indicated by a region of similar (i.e., highly correlated) motion vectors that progress in a trajectory across several frames. A person/object entering or exiting from the edge of the field of view may be indicated by a growing or shrinking region of correlated motion vectors at the corresponding picture boundary. A person/object entering or exiting, e.g., from a doorway, within the field of view will likely be indicated by a growing or shrinking region of motion vectors forming an inward-pointing or outward-pointing radial pattern across a series of frames. Here, too, spatial and temporal patterns of motion vectors and inter/intra block type fields can be used to detect these different situations.

A sequence of frames having a motion vector pattern in which almost all motion vectors are zero motion vectors (i.e., corresponding to essentially no motion across the image) can be used detect the

occurrence of still text, slides, and pictures that occupy the entire video frame, or a general lack of moving objects in the scene.

Motion vector data could also be used to guide noise reduction and edge enhancement processing. If a block is stationary over several frames, these blocks can be averaged together for temporal noise reduction. If motion exists, such averaging would result in unacceptable motion blur. On the other hand, noise reduction could be achieved by averaging after taking motion into consideration. In effect, the motion vector data substitutes for the motion detection in a motion-adaptive noise reduction algorithm. Similarly, motion vector data can be used to implement temporal edge enhancement techniques. In addition, the knowledge of the coarseness of the actual quantization matrices can be used to contrain enhancement processing.

One way to characterize the various patterns of motion vectors would be to have a set of canned motion vector patterns that would be convolved over the decoded motion vector field (e.g., taking vector inner products along the way) to generate a correlation value (e.g., average inner product) that could be compared to a threshold value to determine whether the motion vector field possessed a similar general pattern. Another technique would be to use statistical analysis (e.g., mean and/or standard deviation of motion vector data) over either the entire picture or specific regions to characterize the presence of high-level events in the scene. For example, a set of contiguous blocks having a large mean motion vector and a small standard deviation within an otherwise stationary picture suggests the presence of a moving object within the scene.

<u>Quantized DCT Coefficients</u>

In certain embodiments of the present invention, it may be useful to decode the compressed digital video bitstream sufficiently to extract the quantized DCT coefficients. Depending on the application, it may be further useful to dequantize the extracted quantized DCT coefficients, although even knowledge of the presence of a non-zero quantized DCT coefficients (which emerges from the number and type of run-length/value joint codes and the location of the end of block (EOB) codes) may be enough to provide insight into picture content. In either case, the DCT coefficients can be used to characterize the spatial frequency within each frame as well as the temporal changes in spatial frequency between frames. This information can be used to characterize certain types of picture content. For example, if the DC coefficient of the (B-Y) component DCT block is large in most blocks in the upper third of a frame, it probably corresponds to sky. As another example, in an inter-frame encoded block, low-energy DCT coefficients indicate no substantial change, while high-energy DCT coefficients may indicate a change of shape or texture within the block.

For example, text appearing in an image (either over the entire field of view or just within a region of the image) usually exhibits the combination of having high contrast, being monochromatic, and having many edges. This unusual combination of characteristics may be indicated in the DCT coefficients as

high-energy, high-frequency DCT coefficients in many orientations with few or even no non-zero quantized coefficients in the corresponding U and V blocks.

Similarly, the existence of a raster display in a video sequence (e.g., an image in which an active television or computer display is within the field of view) may be detected by the presence of a temporal beat frequency between the raster display frame rate and the frame rate of the video sequence. Both inter/intra block type fields and DCT level spatio-temporal frequency analysis can contribute to this detection.

The present invention may be implemented as circuit-based processes, including possible implementation on a single integrated circuit. As would be apparent to one skilled in the art, various functions of circuit elements may also be implemented as processing steps in a software program. Such software may be employed in, for example, a digital signal processor, micro-controller, or general-purpose computer.

The present invention can be embodied in the form of methods and apparatuses for practicing those methods. The present invention can also be embodied in the form of program code embodied in tangible media, such as floppy diskettes, CD-ROMs, hard drives, or any other machine-readable storage medium, wherein, when the program code is loaded into and executed by a machine, such as a computer, the machine becomes an apparatus for practicing the invention. The present invention can also be embodied in the form of program code, for example, whether stored in a storage medium, loaded into and/or executed by a machine, or transmitted over some transmission medium or carrier, such as over electrical wiring or cabling, through fiber optics, or via electromagnetic radiation, wherein, when the program code is loaded into and executed by a machine, such as a computer, the machine becomes an apparatus for practicing the invention. When implemented on a general-purpose processor, the program code segments combine with the processor to provide a unique device that operates analogously to specific logic circuits.

It will be further understood that various changes in the details, materials, and arrangements of the parts which have been described and illustrated in order to explain the nature of this invention may be made by those skilled in the art without departing from the principle and scope of the invention as expressed in the following claims.

CLAIMS

What is claimed is:

1. A method for characterizing picture content of a compressed video bitstream, comprising the steps of:

(a) partially decoding the compressed video bitstream to extract particular data for the compressed video bitstream; and

(b) analyzing the extracted particular data to characterize the picture content of the compressed video bitstream.

2. The invention of claim 1, wherein:

the extracted particular data comprises inter/intra block type data; and

step (b) comprises the steps of counting a number of intra-frame blocks per frame and comparing the number of intra-frame blocks to a specified threshold to detect changes in the picture content.

3. The invention of claim 1, wherein the extracted particular data comprises motion vector data.

4. The invention of claim 3, wherein the extracted particular data further comprises inter/intra block type data.

5. The invention of claim 3, wherein step (b) comprises the step of characterizing a pattern in the motion vector data to detect changes in the picture content.

6. The invention of claim 3, wherein step (b) comprises the step of using the motion vector data to guide noise reduction processing.

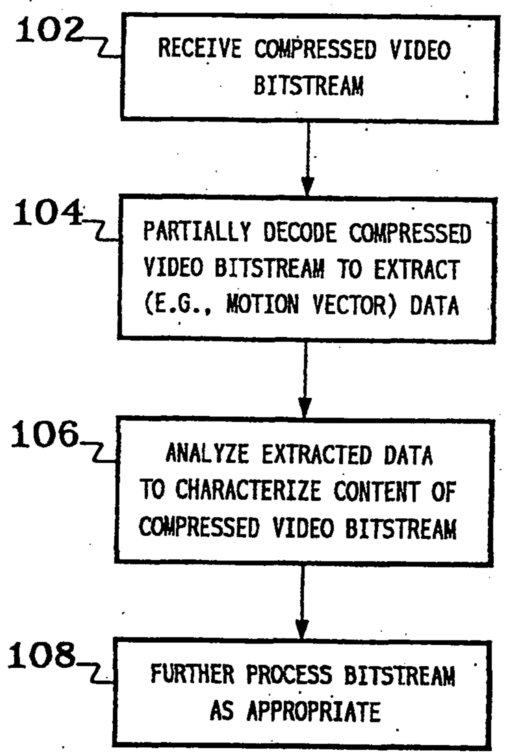
7. The invention of claim 3, wherein step (b) comprises the step of using the motion vector data to guide edge enhancement processing.

8. The invention of claim 1, wherein the extracted particular data comprises DCT coefficient data.

9. The invention of claim 8, wherein the extracted particular data further comprises inter/intra block type data.

10. The invention of claim 8, wherein step (b) comprises the step of characterizing the DCT coefficient data to characterize the picture content.

FIG. 1


102
RECEIVE COMPRESSED VIDEO BITSTREAM
104
PARTIALLY DECODE COMPRESSED VIDEO BITSTREAM TO EXTRACT (E.G., MOTION VECTOR) DATA
106
ANALYZE EXTRACTED DATA TO CHARACTERIZE CONTENT OF COMPRESSED VIDEO BITSTREAM
108
FURTHER PROCESS BITSTREAM AS APPROPRIATE